Inter-physician agreement on the readiness of sick-listed employees to return to work

Schreuder, Jolanda A. H.; Roelen, Corne A. M.; de Boer, Mintje; Brouwer, Sandra; Groothoff, Johan W.

Published in:
Disability and Rehabilitation

DOI:
10.3109/09638288.2012.665125

Publication date:
2012

Disability & Rehabilitation, 2012; 34(21): 1814–1819

ISSN 0963-8288 print/ISSN 1464-5165 online
DOI: 10.3109/09638288.2012.665125

informa
healthcare

RESEARCH PAPER

Inter-physician agreement on the readiness of sick-listed employees to return to work

Jolanda A. H. Schreuder[1], Corné A. M. Roelen[1,2], Mintje de Boer[1], Sandra Brouwer[1] & Johan W. Groothoff[1]

[1]Department of Health Sciences, Community and Occupational Medicine, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands and [2]365/ArboNed Occupational Health Service, Zwolle, The Netherlands

Purpose: **To determine the agreement between occupational physician (OP) ratings of an employee's readiness to return to work (RRTW).** ***Method:*** **Anonymized written vignettes of 132 employees, sick-listed for at least 3 weeks, were reviewed by 5 OPs. The OPs intuitively rated RRTW as the ability (knowledge and skills) and willingness (motivation and confidence) of sick-listed employees to resume work. Inter-OP percentages of agreement were calculated and Cohen's kappas (κ) were determined to correct for agreement by chance.** ***Results:*** **The percentage of agreement between OPs was 57% (range 39–89%) on the ability and 63% (range 48–87%) on the willingness of sick-listed employees to resume work. The mean κ was 0.14 (range from –0.21 to 0.79) for ability and 0.25 (range from –0.11 to 0.74) for willingness. The OP-rating of RRTW of employees sick-listed with mental disorders did not differ from the OP-rating of RRTW of employees with musculoskeletal disorders.** ***Conclusion:*** **The inter-OP agreement on intuitively rated RRTW showed a wide variability, which accentuates the need for instruments to establish an employee's RRTW and for training in giving well founded return to work recommendations.**

Keywords: Sickness absence, readiness to return to work, occupational disability, judgement, between-physician agreement, medical decision-making

Implications for Rehabilitation

- There is little agreement between physicians' intuitive ratings of the readiness of sick-listed employees to return to work.
- This finding accentuates the need for an instrument to establish an employee's readiness to return to work more consistently.
- A reliable assessment of the readiness to return to work is important for motivating sick-listed employees to resume work and for advising managers on how to support and instruct the sick-listed employee.

Introduction

The prevalence of sickness absence is high in many Western societies. In the Netherlands, approximately 56% of women and 49% of men report to have been absent from work due to injuries or illness in the past 12 months, although only 8% of women and 6% of men had been absent longer than 6 weeks [1]. Employees report sick when symptoms and impairments are too serious to continue work, but previous research has shown that other factors impinge on an employee's decision to stay off work when ill [4–7]. For example, employees with nonspecific disorders are uncertain about their symptoms and frequently play it safe by reporting sick to prevent their symptoms from getting worse [8]. Insecurities and uncertainties about poorly-defined symptoms may also adversely affect an employee's readiness to return to work (RRTW).

Readiness to return to work

The RRTW reflects whether or not a sick-listed employee is ready to resume work. The RRTW-model posits that sick-listed employees go through five stages to get ready for work [9]: pre-contemplation (not yet thinking about resuming work), contemplation (considering return to work in the foreseeable future), preparation (actively gathering information for a plan on return to work in the near future), action (putting a return to work plan into action), and maintenance (staying at work). The RRTW-model considers employee-assessed readiness to resume work. To our opinion, it is to be preferred that occupational health providers establish an employee's RRTW so they can appropriately guide sick-listed

Correspondence: Dr. C. Roelen; 365/ArboNed, PO Box 158, 8000 AD Zwolle, The Netherlands. Tel: +31 38 4554700. Fax: + 31 38 4537272.
E-mail: corne.roelen@365.nl
(Accepted February 2012)

employees through the stages of RRTW and recommend work activities.

Hersey and Blanchard's situational leadership theory was developed to train managers in effectively adjusting their leadership style to the readiness level of their employees. The readiness level includes both an employee's ability (i.e. the knowledge and skills) and willingness (i.e. the motivation and confidence) to complete a task [10,11]. This theoretical framework can also be applied to occupational rehabilitation. Employees at the lowest RRTW level lack both the ability and willingness to return to work [11]. "Enthusiastic beginners" are willing to resume work, but lack the knowledge or skills to do so, whereas "cautious performers" have the ability to resume work, but lack the willingness. At the highest RRTW level, "self-reliant achievers" are both able and willing to resume work. The rating of the RRTW level may provide tools for health providers in giving return to work recommendations.

RRTW and return to work recommendations

The assessment of RRTW and the recommendation of work activities depend on a wide range of factors, among others the severity of illness, intensity of symptoms, functional capacity, physical work demands, and the employee's beliefs and expectations. In the Netherlands, occupational physicians (OPs) advise and guide sick-listed employees during the process of return to work. The OP-rating of an employee's RRTW is usually based on the ideas and plans employees have about resuming work and on the OP's intuitions. Intuitive OP-ratings result from implicate a-priori medical knowledge and experience in occupational healthcare and not from structured protocols, procedures or decision-making. Intuitive judgement is a complex issue in itself and one could argue that OPs use time-saving heuristic decisions to maneuver more efficiently through the assessment of RRTW. However, judgements and recommendations associated with return to work have important consequences for sick-listed employees and their employers. Therefore, the assessment of RRTW and the advice about return to work should be based on normative decision-making in which judgements and decisions result from careful comparisons of logical and rational rules, instead of heuristic decision-making [12].

Three previous studies have examined physicians' agreements on return to work recommendations. Rainville et al. distributed written vignettes describing three patients with low back pain among 142 practicing physicians and found a modest reliability (57% agreement) of return to work recommendations [13]. Chibnall et al. also used written vignettes of patients with chronic low back pain, which were scored by 48 internal medicine physicians. They found a high within-physician consistency, but a very low between-physician agreement (correlation 0.11) in the rating of the occupational disability level [14]. Ikezawa et al. composed three case reports providing detailed information on past medical history, present injury, physical examination findings and occupational demands [15]. The diagnoses for the three case reports were fracture, dislocation, and back pain. There was a high percentage of agreement between 36 health providers in giving return to work recommendations for employees with fractures (97.2%) or dislocations (94.4%). However, the agreement on the back pain scenario was modest (55.6%), which was explained by the fact that the etiology of pain and its relationship with disability are more complex.

Purpose of the study

To our knowledge, there are no studies that investigated the reliability of OP-rated RRTW. OPs intuitively assess an employee's RRTW by rules of thumb based on their experience in occupational healthcare. How reliable are these intuitive OP-ratings? This study determined the inter-OP agreement in the intuitive ratings of an employee's RRTW level defined according to Hersey and Blanchard's situational leadership theory.

Methods

Study setting

The study was performed by using information of employees working in a somatic hospital, which contracted OP1 as the company physician. The hospital employees reported sick to their manager when they were too ill to attend work. If a sick-listed employee resumed work within the first 2 weeks of calling in sick, then medical certification of sickness absence was not required. A sick-listed employee visited OP1 in the third week of sickness absence for a medical certification of sickness absence. The consultations of OP1 provided insight into medical factors (symptoms, diagnosis, and treatment), work-related factors (work content, work conditions, work environment), private life (family, leisure time activities, life-events, lifestyle), behavioral factors (coping, personality, self-efficacy, sense of coherence), and attitudes towards return to work (sickness absence values, fear-avoidance beliefs, irrational illness cognitions, ideas about resuming work). This information was recorded in the employees' medical files.

Written vignettes of cases

In 2009, OP1 had consulted 132 hospital employees who had been sick-listed for at least three consecutive weeks. The files of the first consultations with these employees were anonymized and printed in 2010. Besides OP1, four other OPs were asked to review these written vignettes of sick-listed employees. After a brief instruction on Hersey and Blanchard's RRTW levels, all five OPs reviewed the written vignettes on the ability (high–low) and willingness (high–low) of employees to resume work. Ability was rated high if the employee had ideas about resuming work or saw opportunities to accommodate work. Ability was rated low if the employee was not yet thinking about resuming work. Willingness was rated high if the employee was motivated and confident to resume or accommodate work and low if the employee foresaw problems in resuming work or was not confident for example due irrational beliefs or fear-avoidance behavior.

Ethical considerations

In the Netherlands, ethical clearance is not required for studies of anonymous data or records, provided that the researcher

does not carry out any procedures which disclose the identity of the involved individuals [16].

Statistical analysis

The percentage agreement between the ratings of OPs was calculated for all possible 2×2 OP pairs, being: OP1×OP2, OP1×OP3, OP1×OP4, OP1×OP5; OP2×OP3, OP2×OP4, OP2×OP5; OP3×OP4, OP3×OP5, and OP4×OP5. Generally, an agreement of less than 60% is considered poor, 60–80% modest, and >80% good [17]. An important weakness of calculating the percentage of agreement is that it does not take into account the agreement that is expected to occur by chance. The κ-statistic (κ) corrects for the fact that observers sometimes agree or disagree simply by chance and Cohen's κ is the most commonly used statistic to estimate inter-observer reliability [17–21]. Cohen's κ was calculated for the inter-observer reliability of all possible 2×2 OP pairs. The κ statistic has a maximum of 1.00 when agreement is perfect and a value of 0.00 when there is no agreement better than chance. Values of κ = 0.81–1.00 reflect excellent reliability, κ = 0.61–0.80 good reliability, κ = 0.41–0.60 moderate reliability, κ = 0.21–0.40 fair reliability, and κ = 0.00–0.20 poor reliability. A Cohen's κ < 0.00 reflects systematic disagreement.

Inter-OP percentages of agreement and κ statistics were also calculated separately for the OP-rating of RRTW of employees sick-listed with mental symptoms and employees sick-listed with musculoskeletal symptoms. The agreement between the OP-ratings of RRTW in both groups of employees were compared by using the non-parametric Mann-Whitney test concluding significance for $p < 0.05$.

Results

Five OPs (3 women and 2 men) with a mean age of 45.4 (range 38–57) years, who worked as an OP for on average 10.6 (range 8–16) years, independently studied the written vignettes of 132 employees (118 women and 14 men) with a mean age of 45.0 (standard deviation 9.6) years and sick-listed with mental disorders (n = 61), musculoskeletal disorders (n = 50), or other disorders (n = 21), predominantly cardiovascular, gastrointestinal, and neurological disorders. Based on the vignettes, the OPs rated the employees' ability and willingness to return to work.

Overall agreement between OPs

The overall percentage agreement was 57% (range 39–89%) for the OP-rating of ability and 63% (range 48–87%) for the willingness of sick-listed employees to resume work. OP1 had consulted all 132 sick-listed employees and may therefore have had fuller information on them. Excluding OP1 from the analyses yielded similar results with a 61% agreement (range 49–89%) on ability and 66% (range 48–87%) agreement on willingness. The inter-OP reliability is shown in Table I for each possible 2×2 OP-pair. The mean κ for an employee's ability to resume work was 0.14 (range from –0.21 to 0.79) with one OP pair showing good agreement, one pair moderate agreement, and one pair fair agreement. Seven pairs showed poor agreement on the rating of the ability to resume work. The mean κ for willingness was 0.25 (range from –0.11 to 0.74), with four pairs showing good agreement and six pairs poor agreement.

Agreement between OPs on records of employees with mental disorders

Of the 132 employees, 61 (46%) were sick-listed with mental disorders: 51 stress-related disorders and 10 depressive disorders. The inter-OP agreement on the rating of the ability of employees with mental disorders to resume work was 55% (range 36–87%) with a mean κ = 0.05 (range from –0.16 to 0.52). One 2×2 OP-pair showed moderate agreement, one pair fair agreement, and eight pairs poor agreement (Table II). The inter-OP agreement on the rating of the willingness of employees with mental disorders to resume work was 58% (range 38–87%) with a mean κ = 0.18 (range from –0.12 to 0.63). Two OP pairs showed good agreement, one pair moderate agreement, one pair fair agreement, and six pairs poor agreement.

Table I. Overall agreement between occupational physicians (OP) in rating the written medical records of 132 employees

		Ability								Willingness							
		OP2		OP3		OP4		OP5		OP2		OP3		OP4		OP5	
		high	low	high	low	high	low	high	low	high	low	high	low	high	low	high	low
OP1	high	56	14	32	38	27	43	32	38	49	20	35	34	37	32	30	39
	low	22	40	41	21	26	36	39	23	5	58	32	31	40	23	34	39
	κ	*0.45*		*–0.21*		*–0.03*		*–0.17*		*0.62*		*0.00*		*–0.10*		*–0.11*	
OP2	high			42	36	33	45	41	37			30	24	31	23	26	28
	low			31	23	20	34	30	24			37	41	46	32	38	40
	κ			*–0.04*		*0.05*		*–0.03*				*0.08*		*–0.02*		*–0.01*	
OP3	high					40	33	65	8					60	7	57	10
	low					13	46	6	53					17	48	7	58
	κ					*0.32*		*0.79*						*0.64*		*0.74*	
OP4	high							38	15							60	17
	low							33	46							4	51
	κ							*0.28*								*0.68*	

The kappa statistic (κ) is a measure for inter-observer reliability corrected for the agreement expected by chance.

Table II. Agreement between occupational physicians (OP) in rating the written medical records of 61 employees with mental disorders

		Ability								Willingness							
		OP2		OP3		OP4		OP5		OP2		OP3		OP4		OP5	
		high	low	high	low	high	low	high	low	high	low	high	low	high	low	high	low
OP1	high	23	3	19	7	13	13	20	6	20	8	19	9	22	6	17	11
	low	17	18	32	3	19	16	31	4	3	30	22	11	29	4	24	9
	κ	*0.37*		*–0.16*		*–0.04*		*–0.10*		*0.63*		*0.01*		*–0.09*		*–0.12*	
OP2	high			33	7	21	19	33	7			17	6	18	5	15	8
	low			18	3	11	10	18	3			24	14	33	5	26	12
	κ			*–0.04*		*0.00*		*–0.04*				*0.09*		*–0.07*		*–0.03*	
OP3	high					27	24	47	4					39	2	36	5
	low					5	5	4	6					12	8	5	15
	κ					*0.02*		*0.52*						*0.40*		*0.63*	
OP4	high							26	6							38	13
	low							25	4							3	7
	κ							*–0.05*								*0.32*	

The kappa statistic (κ) is a measure for inter-observer reliability corrected for the agreement expected by chance.

Table III. Agreement between occupational physicians (OP) in rating the written medical records of 50 employees with musculoskeletal disorders

		Ability								Willingness							
		OP2		OP3		OP4		OP5		OP2		OP3		OP4		OP5	
		high	low	high	low	high	low	high	low	high	low	high	low	high	low	high	low
OP1	high	25	8	9	24	9	24	8	25	23	7	11	19	10	20	10	20
	low	3	14	4	13	6	11	4	13	1	19	7	13	6	14	6	14
	κ	*0.54*		*0.03*		*–0.06*		*0.01*		*0.68*		*0.02*		*0.03*		*0.03*	
OP2	high			6	22	8	20	4	24			11	13	9	15	9	15
	low			7	15	7	15	8	14			7	19	7	19	7	19
	κ			*–0.10*		*–0.03*		*–0.21*				*0.19*		*0.11*		*0.11*	
OP3	high					10	3	11	2					15	3	16	2
	low					5	32	1	36					1	31	0	32
	κ					*0.60*		*0.84*						*0.82*		*0.91*	
OP4	high							9	6							15	1
	low							3	32							1	33
	κ							*0.55*								*0.91*	

The kappa statistic (κ) is a measure for inter-observer reliability corrected for the agreement expected by chance.

Agreement between OPs on records of employees with musculoskeletal disorders

Of the 132 employees, 50 (38%) were sick-listed with musculoskeletal disorders: 22 employees had arthrosis or spondylosis, 10 injuries, and 6 herniated disks; 12 employees had nonspecific musculoskeletal pain. The inter-OP agreement in the rating of the ability of employees with musculoskeletal disorders to return to work was 59% (range 36–94%) with a mean $\kappa = 0.22$ (range from –0.21 to 0.84). One OP pair showed excellent agreement, three pairs moderate agreement, and six pairs poor agreement on the ability of an employee with musculoskeletal disorders to resume work (Table III). The inter-OP agreement in the rating of the willingness of employees with musculoskeletal disorders to return to work was 68% (range 56–96%) with a mean $\kappa = 0.38$ (range from 0.02 to 0.91). Three OP pairs showed excellent agreement, one pair moderate agreement, and six pairs showed poor agreement on the willingness of an employee with musculoskeletal disorders to resume work.

The inter-OP reliability of the RRTW-rating of employees sick-listed with mental disorders did not differ significantly from reliability in the RRTW-rating of employees sick-listed with musculoskeletal disorders with Mann–Whitney p = 0.27 for ability and p = 0.08 for willingness.

Discussion

The results showed that the agreement between OP-ratings of an employee's RRTW was poor for the ability and modest for the willingness to resume work. The percentages of agreement were in-line with those found for return to work recommendations in back pain patients [13–15]. As in the other studies on return to work recommendations, there was a wide variability in the OP-rating of RRTW, ranging from systematic disagreement to good agreement and sometimes even excellent agreement. It should be acknowledged that excellent inter-OP agreement does not necessarily mean that the ratings accurately reflected the actual RRTW. After all, the inter-OP agreement is a measure for the reliability of RRTW-ratings and not for the validity. For example, the inter-OP agreement may be very high if two OPs are equally wrong about an employee's RRTW. The RRTW-ratings should

be associated with the progress or outcome of occupational rehabilitation to assess their validity.

The variability of OP-ratings may be explained by the multifactorial diversity of occupational disability. It may also be the result of decisional heuristics if OPs use different rules of thumb when recommending work activities. Rainville et al. reported that a physician's appraisal of pain and perception of severity of symptoms accounted for the variability in work recommendations [13]. Chibnall et al. confirmed that physicians were more consistent in their judgement of occupational disability when pain was high. Physical examination and functional disability information did not add to the consistency of physicians' occupational disability judgements[14]. Obviously, the inter-physician variability in judgements on occupational disability and return to work recommendations is associated with physicians' attitudes and beliefs rather than clinical information.

The poor agreement and wide range in the OP-ratings, found in this study, accentuate the poor reliability of intuitively rated RRTW and underline that return to work recommendations should not be solely based on the knowledge and experience of OPs. Standardized instruments have a greater chance of having acceptable consistency and reliability of work-related assessments[22]. Spanjer et al. found a 76% (range 64–88%) agreement between 12 insurance physicians (IPs) when IPs used standardized instruments to record physical and mental work limitations to rate occupational disability[23].

There is an instrument to asses RRTW, but this tool is employee-administered and scores may be biased by an employee's feelings of uncertainty and fear-avoidance beliefs. It is important to develop a physicians' instrument to assess an employee's RRTW so that OPs can give well founded return to work recommendations. Furthermore, knowledge about an employee's RRTW is also important for managers, so they know how to support and instruct sick-listed employees in the return to work process. Employees who lack the ability to resume work need a task-oriented approach, while employees who lack the willingness are best supported by a relationship-oriented leadership style [10].

Limitations of the study

Bias in judgements of functional outcomes usually contributes to higher reliability measures [12]. For example, physicians frequently have images of how patients might appear, a phenomenon known as "representativeness heuristics" [15]. It is unlikely that such representativeness heuristics biased the results of this study, unless the OPs had different ideas of how employees at each RRTW level would appear. Public opinions about patients also bias physicians' judgements of functional outcomes [24]. For example, psychiatric diagnoses elicit stigmatizing responses separate from those directly attributable to symptomatic behavior [25]. Such bias by diagnosis was unlikely, as the OP-ratings of RRTW of employees with mental disorders did not differ from the ratings of employees with musculoskeletal disorders. The OPs rated the written vignettes independent of each other preventing bias by colleague ratings [24]. OP1 and OP2 worked together in a partnership and therefore their ratings may be colleague-biased. Furthermore, OP1 consulted all 132 employees in 2009 and, despite the fact that the study was performed 1 year later and the written vignettes were anonymized, it could not be ruled out that OP1 may have had fuller information on the employees. However, colleague-bias and bias by fuller information were unlikely, because similar inter-OP agreements were obtained after excluding the ratings of OP1 from the analyses. However, bias by availability heuristics due to recent clinical observations or experiences could not be excluded [24].

Another limitation is that the study only included 5 OPs to rate the written vignettes of sick-listed employees. Including more OPs might increase the precision of the κ statistic and reduce its variability. However, including more OPs would not affect the range of the κ statistics. Moreover, the κ statistics were so low that it is to be expected that the inter-OP agreement on the intuitive RRTW-rating of sick-listed employees will remain unsatisfactory, even after including more OPs. The κ statistic corrects for the observed agreement expected by chance, but may be difficult to interpret when data are skewed or the number of observations is low. Lack of variation in the cell fillings may result in a large discrepancy between the percentages of agreement and κ statistics. However, in this study the inter-OP reliability was poor with regard to both the percentages of agreement and κ statistics.

Another limitation of the study is that the rating of RRTW was based on the medical records and only OP1 had consulted the employees. The medical information was recorded in 2009 without having the aims and purpose of this study in 2010 in mind. Possibly, the medical records lacked the data to appropriately assess an employee's RRTW and the agreement between OP-ratings may have been better when all OPs had had the opportunity to consult the sick-listed employees [23]. A study design in which all five OPs consult the same sample of sick-listed employees would better reflect the RRTW-ratings in daily occupational healthcare practice. However, the sparse literature on return to work recommendations was based on written vignettes [12–14,23]. Also, written vignettes of employees sick-listed with mental disorders are used to assess the quality of occupational healthcare by checking whether OPs recommend return to work according to the guidelines of the Dutch Occupational Medicine Association. The present results revealed the uncertainty of relying on written vignettes.

Finally, the RRTW-ratings were not calibrated and the instruction of OPs on the construct of RRTW was brief, because we wanted to investigate the reliability of intuitive RRTW-ratings. Calibration, for example by pre-measurement of agreement on pilot cases, and more extensive training may improve the inter-OP agreement on RRTW of sick-listed employees.

Conclusion

Despite the shortcomings of the study, we conclude that the inter-OP agreement on the intuitive rating of an employee's RRTW was poor and showed a wide variability. This accentuates the need for instruments to structure OPs' consultations and for training of OPs in the assessment of an employee's RRTW.

Acknowledgement

The authors wish to thank Greetje Akker (MD), Jan de Graaf (MD, PhD), Jeep Schreuder (MD) and Bea Rozema (MD) for the time and effort they put in reviewing the written vignettes of 132 sick-listed employees.

***Declaration of Interest*:** The authors report no declarations of interest

References

1. Van den Bossche SNJ, Hupkens CLH, de Ree SJM, Smulders PWG editors. Netherlands Working Conditions Survey 2005: methodology and overall results. Hoofddorp: TNO Work and Employment; 2006.
2. Marmot M, Feeney A, Shipley M, North F, Syme SL. Sickness absence as a measure of health status and functioning: from the UK Whitehall II study. J Epidemiol Community Health 1995;49:124–130.
3. World Health Organisation. Towards a common language for functioning, disability and health. Geneva:WHO; 2002.
4. Kristensen P. Sickness absence and work strain among Danish slaughterhouse workers: an analysis of absence from work regarded as coping behaviour. Social Science and Medicine 1991;32:15-27.
5. Mechanic D. Sociological dimensions of illness behaviour. Social Science and Medicine 1995;41:1207-1216.
6. Petrie KJ, Weinman JA, editors. Perceptions of health and illness: Current research and applications. Amsterdam:Harwood Academic Publishers; 1997.
7. Johansson G, Lundberg I. Adjustment latitude and attendance requirements as determinants of sickness absence or attendance. Empirical tests of the illness flexibility model. Soc Sci Med 2004;58:1857–1868.
8. Hooftman WE, Westerman MJ, van der Beek AJ, Bongers PM, van Mechelen W. What makes men and women with musculoskeletal complaints decide they are too sick to work? Scand J Work Environ Health 2008;34:107–112.
9. Franche RL, Corbière M, Lee H, Breslin FC, Hepburn CG. The Readiness for Return-To-Work (RRTW) scale: development and validation of a self-report staging scale in lost-time claimants with musculoskeletal disorders. J Occup Rehabil 2007;17:450–472.
10. Hersey P, Blanchard KH, Johnson DE, editors. Management of Organizational Behavior: Leading Human Resources. 8th ed. Upper Saddle River: Prentice Hall; 2001.
11. Blanchard KH, editor. Leading at a higher level. Upper Saddle River: Prentice Hall; 2007.
12. Sox HC, Blatt MA, Higgings MC, Marton KI. Medical decision making. Stoneham (MA): Butterworth; 1988
13. Rainville J, Carlson N, Polatin P, Gatchel RJ, Indahl A. Exploration of physicians' recommendations for activities in chronic low back pain. Spine 2000;25:2210–2220.
14. Chibnall JT, Dabney A, Tait RC. Internist judgments of chronic low back pain. Pain Med 2000;1:231–237.
15. Ikezawa Y, Battié MC, Beach J, Gross D. Do clinicians working within the same context make consistent return-to-work recommendations? J Occup Rehabil 2010;20:367–377.
16. Available at: http://www.federa.org/sites/default/files/bijlagen/coreon/code_of_conduct_for_medical_research_1.pdf.Accessed30September, 2011.
17. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159–174.
18. Altman DG, editor Practical statistics for medical research. London: Chapman and Hall; 1991. p 403–409.
19. Rigby AS. Statistical methods in epidemiology. v. Towards an understanding of the κ coefficient. Disabil Rehabil 2000;22:339–344.
20. Kraemer HC, Perikyakoil VS, Noda A. κ coefficients in medical research. Statistics in Medicine 2002;21:2109-2129.
21. Viera AJ, Garrett JM. Understanding interobserver agreement: the κ statistic. Fam Med 2005;37:360–363.
22. Innes E, Straker L. Reliability of work-related assessments. Work 1999;13:107–124.
23. Spanjer J, Krol B, Brouwer S, Groothoff JW. Inter-rater reliability in disability assessment based on a semi-structured interview report. Disabil Rehabil 2008;30:1885–1890.
24. Wolfson AM, Doctor JN, Burns SP. Clinician judgments of functional outcomes: how bias and perceived accuracy affect rating. Arch Phys Med Rehabil 2000;81:1567–1574.
25. Guimon J. The biases of psychiatric diagnosis. British Journal of Psychiatry 1989;154:33-37.